ChinaXiv合作期刊

知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于 SBM-Malmquist 的技术问答社区知识交流效率测度研究

丁楠 曹玮倬 相甍甍
吉林财经大学管理科学与信息工程学院 长春 130117

摘要：[目的/意义]目前在技术问答社区知识交流方面的研究尚不完善，旨在采用数据包络分析法研究技术问答社区知识交流效率，从而发现其不同板块之间存在的差异和变化规律，为改进社区管理和提高知识交流效率提供参考。[方法/过程]在构建知识交流效率评价指标体系的基础上，以“开源中国”社区 15 个热门板块为研究样本，基于 SBM 模型与 Malmquist 指数计算其知识交流的静态和动态效率值并进行深度分析。[结果/结论]“开源中国”社区 15 个热门板块之间的平均知识交流效率良好，但板块间差距较大，且总体知识交流效率处于上升趋势，主要是由于技术进步变化指数呈现 19.50% 的上涨幅度，说明技术振兴是“开源中国”社区知识交流效率上升的主要原因，但出现一定程度管理不当现象并且资源利用率较低。

关键词：技术问答社区 SBM 模型 Malmquist 指数 知识交流效率 开源中国

分类号：G202

引用格式：丁楠，曹玮倬，相甍甍. 基于 SBM-Malmquist 的技术问答社区知识交流效率测度研究 [J/OL]. 知识管理论坛，2022, 7(2): 101-115[引用日期]. http://www.kmf.ac.cn/p/278/.

1 引言

随着网络技术的蓬勃发展，信息交流和传播的方式开始多样化，越来越多的人在论坛和博客上通过发文(帖)和回文(帖)来表达观点、共享信息和获取知识，人际交流突破了时空的局限。用户提出问题后，社区吸引、发动来自各个领域的其他用户解答问题或提出对策，从而围绕一个主题不断地达成共识，形成了类似于现实中社区的网络团体——虚拟社区。近年来，虚拟社区逐渐成为人们通过互联网交流和获得知识的重要媒介，这使一些搜索引擎无法直接检索到的、存储在人脑当中的知识得以展现。专业性的网络论坛作为一类特殊的虚拟社区，是知识交流的重要工具之一，用户一问一答之间，便发生了信息交换和知识交流[1]，例如技术问答社区是把 IT 技术作为论坛的主题，为技术人员提供知识交流的虚拟社区，其鼓励相关用户都能积极参与开放性讨论并分享创作过程，逐渐成为个人或团体开展知识交流和共

作者简介：丁楠，讲师，博士，硕士生导师；曹玮倬，硕士研究生，通信作者，E-mail: cancer_0001@163.com；相甍甍，讲师，博士，硕士生导师。

收稿日期：2021-10-22 发表日期：2022-03-10 本文责任编辑：刘远颖

享的有效平台。近年来，随着 IT 技术的广泛应用，技术问答社区的规模也逐渐扩大 [2]，本文主要对技术问答社区的知识交流效率进行研究，根据实际得到的数据结果，有针对性地提出优化和管理措施，从而增强用户知识交流效果以及创新力度，助力技术问答社区的持续发展。

“开源中国”社区是目前国内最大的开源技术社区 [3]，自 2008 年创办以来，致力于为 IT 开发者提供发现、使用和交流 IT 技术的问答平台，本文选取“开源中国”社区中最热门的 15 个标签板块作为用户知识交流的数据来源。首先，在梳理和总结已有文献成果的基础上，为测算技术问答社区知识交流效率投入产出值构建相应的评价指标体系；其次，采用数据包络分析方法中的 SBM（Slack Based Measure）模型和 Malmquist 指数计算知识交流静态和动态效率值；再次，根据测算结果，对影响技术问答社区知识交流效率的主要因素进行分析；最后，根据所得出的结论，提出合理化建议，以期为促进技术问答平台知识交流，提高其知识创新水平提供参考意见。

2 相关研究

2.1 问答社区

目前，国内外学者对于问答社区的研究主要关注用户信息行为、用户问答质量以及社区间的对比等方面。关于用户信息行为的研究主要集中于知识贡献方面，例如，刘雨农等以知乎网“疫苗”话题为样本，研究虚拟知识社区的社会网络结构特征及影响因素，发现其用户关系是多层次、块状分布的，存在知识生产者和需求者两个中心，其中核心用户比边缘用户更倾向于贡献知识 [4]。Z. Liu 等研究发现知识贡献行为是用户在问答社区中获得所需信息的前提条件，其中用户的参与、兴趣和关联性等因素对用户的回答意愿有重要的影响 [5]。J. H. Jin 等研究发现用户的自我呈现、其他成员的认知和社会学习机会对知识贡献行为都有正向影响 [6]。关于问答质量方面的研究主要以影响因素的特征为主，例如，沈洪洲等以知乎社区为研究对象，总结了 10 个对回答内容质量具有潜在影响的基本特征，回答内容的标注数量越多，重点内容及结构就越清晰，回答内容质量越高，同时，用户更易认同情感倾向积极的回答 [7]。王伟等以知乎社区为研究对象，着重挖掘了答案及其编写者的特征，并通过逻辑回归、支持向量机和随机森林 3 种方法构建特征模型，对答案质量进行全面预测 [8]。关于社区间对比方面的研究主要集中在中外社区运营视角的功能建设上，例如，李丹选取 Quora 和知乎进行对比，着重从产品性能、运营管理和用户特征 3 个层面展开分析，结果发现 Quora 拥有十分精准的功能定位、更简洁的功能区划分，相比知乎具有更多优势和成功经验 [9]。万莉通过测度小木虫和人大经济论坛 8 门学科的知识交流效率，发现人大经济论坛知识交流效率和规模效率低于小木虫论坛，纯技术效率高于小木虫论坛 [10]。

通过现有研究成果可以发现，国内外学者关于问答社区的研究主要集中于信息行为中知识贡献的研究，少数学者对问答质量的影响因素展开了研究，而对于社区间对比的研究较为匮乏且时间较为久远，并未涉及近年来的发展状态。其中，知识贡献主要涵盖了提出问题和浏览信息 [11]，而提问量与浏览量即可代表问答社区的活跃度 [2]，因此，本研究把问答社区中研究相对成熟的知识贡献作为指标体系，把能够反映主题使用情况的热门精选板块作为决策单元，以此为基础利用数据包络分析法对技术问答社区知识交流效率的测度展开研究，以期丰富问答社区的研究。

2.2 知识交流效率

目前，国内外学者尚未就问答社区知识交流效率展开研究，由庞建刚等 [12] 对虚拟学术社区的定义可知，问答社区可以被视为虚拟学术社区的一个子类。当前，对虚拟社区知识交流效率的研究主要集中于效率测度与影响因素两个方面。对于效率测度方面的研究主要基于数据包络分析法，例如，宗乾进等在 2014 年首次提出了知识交流效率评价指标体系，以科学网

知识管理论坛，2022（2）:101-115
DOI: 10.13266/j.issn.2095-5472.2022.009

博客 8 个学科为例，采用数据包络分析法（Data Envelopment Analysis，DEA）方法进行了静态分析[13]；晋升在 DEA 方法的基础上进一步拓展了 Malmquist 指数方法对小木虫论坛 12 个板块的知识交流效率进行研究[14]；杨瑞仙等以在线健康社区“丁香园论坛”为对象，剔除外生因素后基于三阶段 DEA 模型对其中 5 个板块进行了研究[15]。也有少数学者采用其他研究方法，例如，庞建刚等选取经管之家论坛的 3 个区域为对象，采用随机前沿分析法（Stochastic Frontier Analysis，SFA）及 Kernel 方法对知识交流效率及其动态演化进行研究[12]；胡德华等以 4 个学术虚拟社区的 16 个版块为样本，运用遗传投影寻踪算法进行知识交流效率的综合评估[16]。关于影响因素的研究主要通过结合效率测度进行综合分析，例如，吴佳玲采用 Super-SBM 方法及 Tobit 模型对小木虫论坛的知识交流效率及影响因素进行了分析[17]；袁永旭等以丁香园论坛为研究对象，运用 SBM（Slack Based Measure）模型和 Tobit 模型进行知识交流效率和影响因素的研究[18]。

综上，对于虚拟社区知识交流效率的研究以效率测度为主，且大多具有鲜明的学科特色，例如理工类学科的小木虫论坛、经管类学科的经管之家论坛和医学类的丁香园论坛等，而针对近年来活跃高的技术问答社区的研究较少。同时，在运用数据包络分析方法进行知识交流效率测度时，普遍采用 CCR（Charnes & Cooper & Rhodes）模型以及 BCC（Banker & Chames & Cooper）模型，对于 SBM 模型结合 Malmquist 指数的运用较少。针对此现状，本文基于 SBM 模型和 Malmquist 指数相结合的评价方法，对技术问答社区知识交流进行测度，并提出合理化建议。

3 研究设计

3.1 数据来源

本文选择“开源中国”社区中最热门的 15 个标签板块作为决策单元（Decision Making Units，DMU），其中包括“Java”“Android”和“PHP”等，于 2021 年 1 月借助“八爪鱼采集器”工具，分别对 2018-2020 年 15 个板块的数据进行相应指标体系数据的收集，并对采集到的发帖、回帖数据以及时间、投票和作者进行整理统计，以反映实际的知识交流情况。

3.2 研究方法

3.2.1 SBM 模型

DEA 模型于 1978 年被 A. Charnes、W. W. Cooper 和 E. Rhodes 提出，是一种用于评价多投入、多产出指标距离生产前沿面有效性的非参数方法[19]。传统的 DEA 模型为规模报酬不变的 CCR 模型和规模报酬可变的 BCC 模型，自问世便被广泛应用。但传统的 DEA 模型属于“径向”和“角度”的双度量[20]，“径向”要求投入与产出同比例变化，“角度”通常只能考虑一个投入或产出的导向角度，忽略了 DMU 中松弛变量引起计算结果偏差问题[21]。经过学者们多年不断探究，在传统模型的基础上加入更多的限定条件，不断衍生出超效率 DEA、SBM 以及 EBM 等模型。本文使用 K. Tone[22] 于 2001 年提出的在目标函数中引入松弛变量的 SBM（Slack Based Measure）模型，与传统模型相比，其规避了径向和角度问题的影响，且充分考虑了投入和产出的松弛度，使所得相对效率值介于 [0,1]，提升了效率评价的可信度和准确度，更偏重于知识交流效率最大化的考虑[19]。故本文运用 SBM 模型对技术问答社区知识交流效率进行测算，能够对技术问答社区的投入和产出松弛变量进一步优化，即将最小化的松弛变量作为目标函数，用以表明技术问答社区的知识投入冗余和产出不足越少，知识交流效率值则越高。

SBM 模型如公式（1）所示：

$$\rho^* = \min \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{i0}}}{1 + \frac{1}{S_1 + S_2}\left(\sum_{r=1}^{S_1}\frac{s_r^g}{y_{r0}^g} + \sum_{r=1}^{S_2}\frac{s_r^b}{y_{r0}^b}\right)}$$

$$\begin{cases} s.t.x_0 = X\lambda + s^- \\ y_0^g = Y^g\lambda - s^g \\ y_0^b = Y^g\lambda + s^b \\ \lambda \geq 0, s^- \geq 0, s^g \geq 0, s^b \geq 0 \end{cases}$$

公式（1）

其中，ρ^* 代表 DMU 的相对效率值，S_1 和 S_2 为期望及非期望产出要素个数；x_0、y_0 分别代表 DMU 的投入和产出向量；X、Y 分别代表 DMU 的投入和产出矩阵。当且仅当 $\rho^*=1$ 时，DMU 有效；当 $0\leq\rho^*<1$ 时，DMU 无效。

3.2.2 Malmquist 指数

Malmquist指数由瑞典经济学家 S. Malmquist 在 1953 年提出，随后 R. Färe 等将其与数据包络分析法相结合，用来测算生产和经营效率等问题，使其在金融、医疗和工业等领域应用广泛 [14]。Malmquist指数是通过对比决策单元在 t 时期和 t+1 时期的距离函数比值来判定其效率变化情况和发展趋势，弥补了传统 DEA 无法测算动态效率的缺陷 [23]。故本文运用 Malmquist 指数考量技术问答社区知识交流全要素生产率的演化趋势，其能够与数据包络分析法充分结合，进而对技术问答社区知识交流的动态效率进行测算，以便更好地分析社区的变化趋势。

Malmquist 指数如公式 (2) 所示：

$$M(x^{t+1},y^{t+1},x^t,y^t)=$$

$$\left[\frac{D_C^t(x^{t+1},y^{t+1})}{D_C^t(x^t,y^t)}\times\frac{D_C^{t+1}(x^{t+1},y^{t+1})}{D_C^{t+1}(x^t,y^t)}\right]^{\frac{1}{2}}=$$

$$effch\times techch=pech\times sech\times techch=tfpch \quad \text{公式（2）}$$

其中，x^t 和 x^{t+1} 分别表示 t 和 t+1 时期研发投入量；y^t 和 y^{t+1} 分别表示 t 和 t+1 时期研发产出量；tfpch、effch、techch、pech 和 sech 分别代表全要素生产率、技术效率变动、技术进步、纯技术效率变动和规模报酬变动指数。若 $tfpch>1$，则表示全要素生产率有所提高，整体呈上升趋势；$tfpch=1$，则全要素生产率不变，整体趋势稳定；$tfpch<1$，则全要素生产率降低，整体呈退步状态。

3.3 指标体系构建

技术问答社区是专门进行技术交流的问答社区，用户通过网络便可以实现知识共享、交流和创新。在技术问答社区中，知识提供者 [24] 通过发帖将隐性知识外化提供给使用者，知识使用者通过浏览来获取所需信息，同时可以结合自身知识储备使知识进一步外化，知识反馈者通过投票、评论等行为对知识提供者进行反馈，整个过程实现了知识共享、交流和创新 [10]。本文基于万莉 [10] 的研究成果，并结合技术问答社区知识交流的特征构建技术问答社区知识交流效率的指标体系，如表 1 所示：

表 1 技术问答社区知识交流效率评价指标体系

指标类型	指标名称	指标说明
投入指标	X1（用户数量）	反映知识交流中人员投入程度
	X2（发帖数量）	反映知识交流中精力投入程度
	X3（讨论时间）	反映知识交流中时间投入程度
产出指标	Y1（浏览数量）	反映知识在用户间的传播广度
	Y2（回帖数量）	反映知识在用户间的交流热度
	Y3（投票数量）	反映知识在用户间的交流质量

（1）投入指标 X1——用户数量。用户是技术问答社区开展知识交流的主体，知识的提供和传播都依赖于这一主体。本文统计的用户数量为实际参与到知识交流中的人员数量，即各板块参与发帖和回帖的人员数量，不包括仅注册或投票的人员，本文认为这两类用户不能视作知识交流投入。

（2）投入指标 X2——发帖数量。用户依据所探讨的主题选择相应板块进行发帖，发帖数量充分说明用户在这一领域知识交流精力的投入程度。

（3）投入指标 X3——讨论时间。即从发帖到采集时获取到的最后回帖的时间跨度，衡量知识交流的持续时间，反映了用户在整个知识交流过程中的时间投入情况。

（4）产出指标 Y_1——浏览数量。即各个板块中每篇帖子截止到统计时间的浏览数量，直观反映出用户对帖子主题的关注程度，帖子的浏览数量越大，信息传播范围越广泛。

（5）产出指标 Y_2——回复数量。即用户对感兴趣的帖子进行相关内容的回复数量，其中一个用户针对不同帖子进行回复按帖子数量计数，针对同一帖子多次回复按 1 次计数。帖子

知识管理论坛，2022（2）:101-115
DOI: 10.13266/j.issn.2095-5472.2022.009

的持续回复是将主题进一步内化和吸收的过程，直观反映了话题热度，回帖数量越多说明知识交流的程度越高，传播的影响力越大。

（6）产出指标 Y_3——投票数量。即用户针对答案的准确与否投赞成票和反对票，每人每帖限投一票，票数的高低与帖子的优劣程度息息相关，有助于用户快速甄别更优质的帖子，是技术问答社区知识交流效率评价的核心指标。

4 结果分析与讨论

4.1 描述性统计分析

本文通过对“开源中国”社区 15 个热门标签板块的数据依指标体系进行整理与统计，结果如表 2 所示：

表 2 2018-2020 年“开源中国”社区 15 个板块的投入及产出数据

板块	统计	投入指标			产出指标		
		X_1用户数量/人	X_2发帖数量/篇	X_3讨论时间/天	Y_1浏览数量/次	Y_2回帖数量/篇	Y_3投票数量/票
Java	最小值	1 029	502	5 733	405 586	1 920	460
	最大值	1 718	849	42 599	1 396 028	3 129	1 019
	均值	1 471.67	662.67	23 846.33	939 285.33	2 691.67	760.00
	标准差	313.68	142.82	15 057.27	407 991.49	547.27	230.04
Android	最小值	67	44	273	17 398	74	16
	最大值	195	81	4 430	89 414	220	56
	均值	115.33	60.00	1 697.00	46 852.00	124.67	40.00
	标准差	56.76	15.51	1 933.09	30 826.84	67.46	17.28
PHP	最小值	147	42	910	38 027	188	12
	最大值	333	105	7 742	160 703	497	39
	均值	267.33	84.00	3 926.00	90 467.00	373.33	21.33
	标准差	85.21	29.70	2 845.94	51 638.96	133.47	12.50
MySQL	最小值	350	130	809	74 783	434	55
	最大值	646	239	11 995	371 016	930	345
	均值	510.33	190.67	6 208.33	197 203.33	673.33	167.00
	标准差	122.09	45.35	4 574.87	126 278.47	202.86	127.26
Spring	最小值	287	116	1 733	72 563	318	86
	最大值	774	321	26 355	894 298	1 050	149
	均值	538.33	234.00	12 516.00	445 239.33	687.33	124.00
	标准差	199.13	86.51	10 281.54	339 791.90	298.87	27.31
JFinal	最小值	35	22	22	4 276	30	0
	最大值	143	69	976	37 364	145	55
	均值	92.00	51.33	512.67	23 544.00	102.33	22.00
	标准差	44.29	20.89	389.95	14 046.70	51.42	23.76
Python	最小值	172	81	639	46 351	168	15
	最大值	465	217	6 079	353 306	762	275
	均值	340.33	164.33	3 230.33	198 818.67	463.67	146.33
	标准差	123.54	59.61	2 228.31	125 321.99	242.51	106.16

板块	统计	投入指标			产出指标		
		X_1用户数量/人	X_2发帖数量/篇	X_3讨论时间/天	Y_1浏览数量/次	Y_2回帖数量/篇	Y_3投票数量/票
Eclipse	最小值	72	25	529	8 288	56	1
	最大值	317	125	4 181	135 475	349	26
	均值	204.33	82.67	2 268.67	79 721.00	222.33	16.00
	标准差	100.98	42.24	1 495.91	53 094.30	122.86	10.80
Linux	最小值	128	44	503	44 793	111	27
	最大值	287	120	1 506	140 655	258	181
	均值	194.33	81.67	994.00	92 765.00	196.67	117.33
	标准差	67.53	31.03	409.74	39 135.54	62.43	65.64
jQuery	最小值	20	6	57	1 844	16	2
	最大值	127	44	2 076	54 148	155	13
	均值	86.00	31.00	1 040.33	30 889.00	93.67	7.33
	标准差	47.12	17.68	825.08	21 741.44	57.91	4.50
Apache Tomcat	最小值	106	31	206	31 164	137	19
	最大值	314	110	4 936	264 674	330	71
	均值	217.67	78.33	2 609.67	122 596.33	235.33	47.00
	标准差	85.61	34.10	1 931.79	101 834.64	78.83	21.42
Ubuntu	最小值	18	12	41	2 580	13	0
	最大值	129	55	987	62 506	103	8
	均值	63.67	30.33	389.33	28 054.33	49.33	3.00
	标准差	47.40	18.12	424.54	25 274.83	38.73	6.16
CentOS	最小值	53	27	156	12 345	52	8
	最大值	166	78	3 551	88 933	161	46
	均值	112.33	53.00	1 696.67	51 692.67	100.67	21.00
	标准差	46.31	20.83	1 403.64	31 302.41	45.26	17.68
Android SDK	最小值	67	44	273	17 398	74	16
	最大值	195	81	4 430	89 414	220	56
	均值	115.33	57.00	1 697.00	46 852.00	124.67	40.00
	标准差	56.76	16.99	1 933.09	30 826.84	67.46	17.28
Apache ECharts	最小值	44	33	36	11 890	35	0
	最大值	105	82	2 994	136 000	100	72
	均值	74.33	51.33	1 896.33	73 421.00	66.00	25.33
	标准差	24.90	21.82	1 322.54	50 673.11	26.62	33.04

如表 2 所示，不同板块间指标差异较大，总的来说，投入指标数量越多其产出指标数量也相应越多，如 Java 板块 3 个投入指标数量均为最高，其 3 个产出指标数量也最高；用户数量与发帖数量基本呈正比关系，如 Java 板块用户与发帖量均为最高，而用户量较少的 jQuery 和 Ubuntu 板块发帖量也较少；发帖量越多其讨论时间也越长，如 Java 板块发帖量的均值高达

知识管理论坛，2022（2）:101-115
DOI: 10.13266/j.issn.2095-5472.2022.009

23 846.33 篇，而 Ubuntu 板块仅为 389.33 篇。

4.2 基于 SBM 模型分析知识交流静态效率

本文基于 MaxDEA PRO 6.3 件中的 SBM 无导向模型对“开源中国”社区 15 个热门板块的投入和产出数据进行测算，分别得到综合技术效率、纯技术效率和规模效率。综合技术效率是衡量现有投入资源是否达到最优产出状态[25]，测量技术问答社区中知识交流的整体管理水平；纯技术效率是指投入规模一致时各板块最佳的投入产出效果，由此评估技术问答社区在技术和管理等方面的客观状况；规模效率为综合技术效率与纯技术效率之比[26]，是指技术因素保持不变时投入指标的增长所导致的效率变化，反映了技术问答社区的资源配置状况。

根据表 3 所示，“开源中国”社区 15 个热门板块的平均综合技术效率指数介于 [0.652,1]，平均值为 0.841，处于良好水平。2018-2020 年间，每年分别有 6 个、9 个、7 个板块知识交流效率达到 DEA 有效（综合效率指数为 1），而各板块知识交流综合技术效率变化趋势各异，本文将其划分为以下 4 个类型：①成熟型，包括 Java、JFinal 和 Apache Echarts 板 块。 这 3 个板块每年均达到 DEA 有效，说明这些板块整体管理水平达最优状态，投入与产出要素的组合与规模合理，目前正处在生产前沿面上，知识交流效率达到理想效果。②成长型，包括 PHP、Linux 和 CentOS 板块。2018 年这 3 个板块综合技术效率值均较低，但随着时间的变化不断发展，均逐步达到 DEA 有效状态。③波动型，包括 Android、MySQL、Spring、Eclipse、jQuery、Apache Tomcat 和 Android SDK 板块。这 7 个板块的综合技术效率发展趋势不稳定，均呈现出起伏状态，其中 MySQL 和 Apache Tomcat 板块在 2019 年出现最低值，表现为开口向上的具有凹折点的折线，其余 5 个板块在 2019 年均出现最高值，表现为开口向下的具有凸折点的折线，且 Spring、jQuery 和 Android SDK 3 个板块在 2019 年达 DEA 有效状态。④衰退型，包括 Python 和 Ubuntu 板块。这两个板块均在 2018 年达 DEA 有效状态，且 Ubuntu 板块在 2019 年也达 DEA 有效状态，但 2020 年二者均出现大幅下降趋势，说明其要积极寻找原因并借鉴成熟型板块的管理制度与发展模式，对投入产出要素的资源配置进行优化。

表 3 “开源中国”社区 15 个热门板块的知识交流综合技术效率

板块	2018年	2019年	2020年	均值
Java	1	1	1	1
Android	0.619	0.945	0.548	0.704
PHP	0.908	1	1	0.969
MySQL	1	0.795	0.840	0.878
Spring	0.832	1	0.585	0.806
JFinal	1	1	1	1
Python	1	0.966	0.580	0.849
Eclipse	0.711	0.823	0.423	0.652
Linux	0.719	1	1	0.906
jQuery	0.598	1	0.548	0.715
Apache Tomcat	0.879	0.638	1	0.839
Ubuntu	1	1	0.427	0.809
CentOS	0.549	0.732	1	0.760
Android SDK	0.622	1	0.548	0.723
Apache ECharts	1	1	1	1
均值	0.829	0.927	0.767	0.841

除此之外，从各板块知识交流综合技术效率三年均值来看，无效单元占比高达 80.00%，即大部分板块未达到理想状态，顶尖水平与下游水平板块数量相差较大，说明各板块知识交流的差距较大。其中，Android 板块和 Eclipse 板块三年来效率值均未达到有效，Eclipse 板块效率值最低，仅为 0.652，说明这两个板块在投入、产出以及管理上均存在较大问题，是未来关注的重点对象。

由图 1 可知，在 2018 年各板块效率均值为 0.829，存在 9 个板块效率值高于均值；2019 年各板块效率均值达 0.927，相较于 2018 年出现显著提升，存在 11 个板块效率值高于均值；

2020 年各板块效率均值达 0.767，出现大幅度下降趋势，存在 8 个板块效率值高于均值。15 个板块知识交流效率均值呈现一定的波动变化，且在 2020 年最为明显，Eclipse 板块出现 0.423 的极低值，说明 2020 年由于疫情等重大事件造成社区内各板块的知识交流效率差异较大。

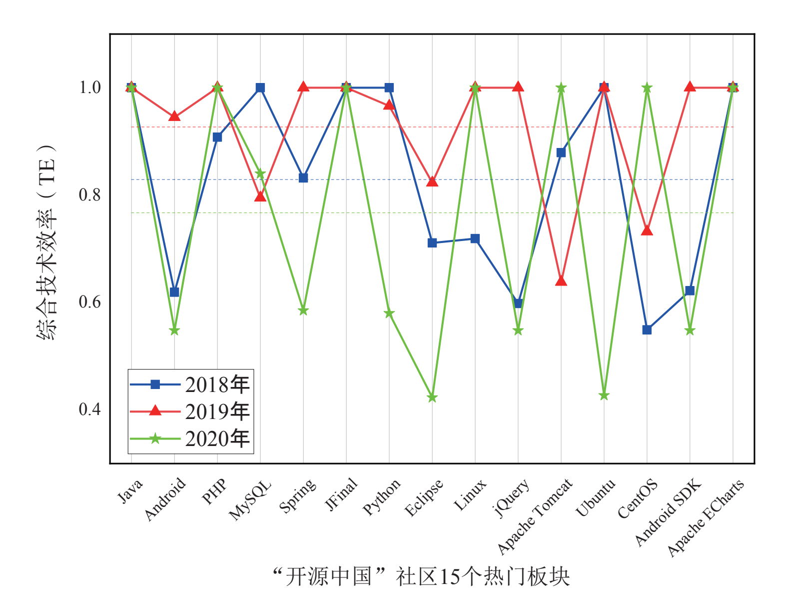


图 1 “开源中国”15 个热门板块综合技术效率变化

根据表 4 可知，导致单元无效的原因可以分为 3 类：①板块的纯技术效率有效而规模效率无效，如 2018 年 PHP、Spring、Linux 和 Apache Tomcat 板块，2019 年 Python 板块以及 2020 年 MySQL、jQuery 和 Ubuntu 板块的纯技术效率均达 1，而规模效率无效且范围处于 [0.427，0.966]，板块间无效值差距较大，说明此类板块现有的技术与管理水平相对完善，在规模保持不变的前提下，充分利用了技术投入来保证知识交流效率的有效，但存在由于规模结构不合理产生的投入冗余现象，导致资源无法充分利用，且板块间实际规模与理想规模间存在不同程度的差距，只有对规模结构进行合理改善，才能有效提升综合技术效率，从而达到 DEA 有效状态；②板块规模效率有效而纯技术效率无效，目前只有 Android 板块在 2019 年规模效率达 1，而纯技术效率无效，说明其规模结构较为合理，资源配置能力较高，对于投入要素的利用较充分且产出效率较高，但当前的技术和管理水平较差，是制约综合技术效率达有效的关键所在，该板块需要强化内部管理，引入较为先进的技术和设备，建立一定的技术壁垒，从而提高纯技术效率；③板块的纯技术效率和规模效率均无效，在 2018-2020 年间每年的板块数分别占比 55.56%、66.67% 和 62.50%，说明此类板块资源配置与技术管理同时出现问题，二者均需加强。在第三类板块中，只有 2018 年 jQuery 板块出现纯技术效率较高，而规模效率较低情况，其余板块均为纯技术效率较低，规模效率较高，且规模效率范围处于 [0.827，0.986]，无效值较为集中且与有效值差距较小，说明其规模效率方面易改进，可以完善投入产出的合理配置，进行社区的精耕细作，从而实现知识交流综合技术效率有效。而纯技术效率方面的改进空间较大，社区内制度和技术管理水平均需要加强，同时应采取积极措施增加用户数量及发帖数量的投入，减少不必要产出，并提升发帖质量，促进投入要素的有效利用，以不断提升纯技术效率。值得一提的是，Android 和 Eclipse 板块三年来从未达到过有效水平。

知识管理论坛，2022（2）:101-115
DOI: 10.13266/j.issn.2095-5472.2022.009

表 4 “开源中国”社区 DEA 无效板块的综合技术效率分解

年份	板块	纯技术效率	规模效率
2018	Android	0.735	0.843
	PHP	1	0.908
	Spring	1	0.832
	Eclipse	0.757	0.939
	Linux	1	0.719
	jQuery	0.895	0.669
	Apache Tomcat	1	0.879
	CentOS	0.664	0.827
	Android SDK	0.738	0.843
2019	Android	0.945	1
	MySQL	0.812	0.979
	Python	1	0.966
	Eclipse	0.847	0.972
	Apache Tomcat	0.693	0.919
	CentOS	0.766	0.957
2020	Android	0.589	0.932
	MySQL	1	0.840
	Spring	0.614	0.953
	Python	0.589	0.986
	Eclipse	0.481	0.878
	jQuery	1	0.548
	Ubuntu	1	0.427
	Android SDK	0.589	0.932

可以发现，“开源中国”社区 DEA 无效板块普遍存在规模效率较低的问题，这说明在知识交流过程中，忽视了规模的扩大，在今后的发展中应多加关注知识投入和产出的合理配置。此外，虽多个板块出现纯技术效率有效情况，但相较于规模效率值所处范围，社区内板块的技术管理水平较高，但两极化现象依旧严峻，制约着综合技术效率的提升，未来应不断加强技术管理水平和制度的约束，建立起能够激发用户参与的互惠政策，促使投入资源得到合理利用，提高知识交流的质量。

4.3 基于 Malmquist 指数分析知识交流动态效率

上文基于 SBM 模型对“开源中国”社区知识交流效率的分析是时间点上的静态分析，为了进一步探究知识交流效率变动的内在因素和趋势，本文根据“开源中国”社区 15 个热门板块在 2018-2020 年间的投入与产出数据，采用 MaxDEA8 软件测算出其全要素生产率指数（tfpch）、技术效率变化指数（effch）和技术进步变化指数（techch）[27]，其中 effch 衡量 DMU 相对于生产前沿面的接近程度，用来反映社区内各板块对现有资源的有效利用能力，可由纯技术效率（pech）和规模效率（sech）决定[28]，techch 代表各个时期最佳效率值的变化规律，用来衡量社区内各板块的技术创新是否随生产前沿变化，上述指标可以直观反映出各板块知识交流动态效率的客观情况。

表 5 为“开源中国”社区 2018-2020 年各生产率指数变动情况。从全要素生产率来看，整个阶段均值为 1.156，年平均增长率为 15.60%，且每个周期指数均大于 1，说明知识交流效率处于上升趋势；从技术效率来看，整个阶段均值为 0.967，说明其总体技术效率变化呈现下降趋势，主要由于 2019-2020 年下降了 25.90%；从技术进步来看，整个阶段均值为 1.195，上升了 19.50%，说明总体技术进步变化指数呈现上涨趋势，主要得益于 2019-2020 年上涨了 41.40%，这表明技术振兴是“开源中国”社区知识交流效率上升的主要原因；从纯技术效率和规模效率来看，均处于下降趋势，平均降幅为 1.30% 和 2.00%，二者在一定程度上限制了技术效率的改善，这与表 3 中所得结果相符。

表 5 “开源中国”社区 2018-2020 年 Malmquist 指数

年份	effch	techch	pech	sech	tfpch
2018-2019	1.105	1.006	1.046	1.057	1.112
2019-2020	0.846	1.420	0.931	0.908	1.201
均值	0.967	1.195	0.987	0.980	1.156

综合前面分析发现，“开源中国”社区知识交流效率良好且 2018-2020 年间全要素生产率处于上升趋势，为探索其深层次原因，下文分别对 15 个热门板块的全时段和分时段 Malmquist 指数进行分析。

表 6 为“开源中国”社区 15 个热门板块全时段 Malmquist 指数情况，约 66.67% 板块的全要素生产率呈现增长态势，15 个板块平均增幅为 15.60%，其中技术效率变化指数平均降幅 3.30%，但技术进步变化指数平均增幅 19.50%，从而拉高了全要素生产率。从图 2 可以明显看出“开源中国”社区全要素生产率增长的动力来源是技术进步，这也得益于近年来我国 IT 技术产业技术水平的显著提升。15 个板块中表现突出的是 JFinal 板块，全要素生效率增幅高达 107.8%，高增长态势主要归因于技术进步增幅高达 107.8%。技术效率呈现下降趋势，即使是呈现全要素生产效率上升状态的板块中也有约 33.33% 出现了技术效率下降的情况，说明技术产业各种资源的协调不够优化，现有的技术潜能也没有得到最大程度的释放，管理水平有待进一步提高。全要素生产率呈下降趋势的 33.3% 板块中，Java 和 Apache Echarts 主要是技术进步出现下降，Python 和 Eclipse 为技术效率退步，而 Spring 技术进步和技术效率均呈现降态。由图 3 可以发现，虽然纯技术效率和规模效率对技术效率存在绝对影响作用，但绝大多数技术效率呈现下降的板块，其规模效率也同时呈现下降趋势，而纯技术效率却没有呈现类似的变化情况，说明应加大对于现有资源结构和规模的整合，提高创新水平，将具有创新性的投入应用于研发关键技术和核心信息，提高资源配置状况，从而推动其各板块技术效率进步。

表 6 “开源中国”社区各板块全时段 Malmquist 指数

板块	effch	techch	pech	sech	tfpch
Java	1.000	0.989	1.000	1.000	0.989
Android	1.009	1.637	1.026	0.984	1.652
PHP	1.002	1.050	1.000	1.002	1.052
MySQL	0.958	1.187	1.000	0.958	1.138
Spring	0.902	0.791	0.845	1.067	0.714
JFinal	1.000	2.078	1.000	1.000	2.078
Python	0.801	1.141	0.804	0.996	0.913
Eclipse	0.839	1.043	0.848	0.989	0.875
Linux	1.073	1.447	1.000	1.073	1.553
jQuery	0.983	1.172	1.070	0.918	1.152
Apache Tomcat	1.004	1.092	1.000	1.004	1.097
Ubuntu	0.733	1.595	1.000	0.733	1.168
CentOS	1.316	0.936	1.270	1.036	1.232
Android SDK	1.007	1.639	1.023	0.985	1.651
Apache ECharts	1.000	0.813	1.000	1.000	0.813
均值	0.967	1.195	0.987	0.980	1.156

知识管理论坛，2022（2）:101-115
DOI: 10.13266/j.issn.2095-5472.2022.009

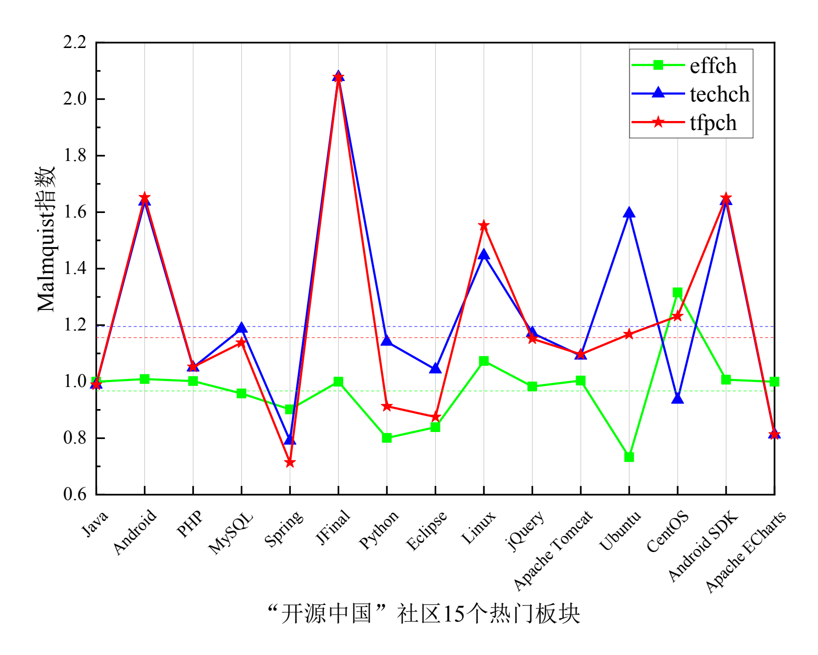


图 2 “开源中国”社区各板块全要素生产率分解情况

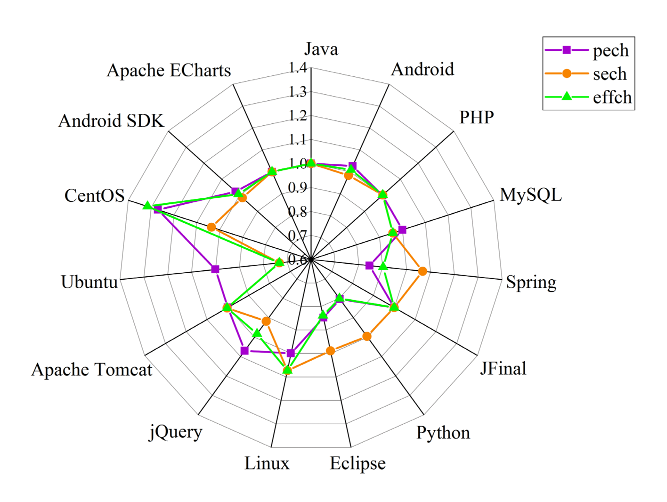


图 3 “开源中国”社区各板块技术效率分解雷达图

表 7 为“开源中国”社区分时段各板块 Malmquist 指数动态变动情况，2018-2019 年间，53.33% 的板块全要素生产率呈现增长态势，其中半数以上涨幅超过 50%，15 个板块平均涨幅 11.20%，并且观察到其技术效率与技术进步均呈同步增涨态势，说明此阶段管理得当，技术水平发展良好，15 个板块中以 Android 板块最为突出，全要素生产率高达 2.887。2019-2020 年间，53.33% 的板块全要素生产率的增长态势，但技术效率呈下降趋势，技术进步呈上升趋势，均变化明显，技术效率的退步是由于纯技术效率和规模效率分别出现 6.90% 和 9.20% 的退步所致，说明虽然在此期间技术发展水平持续升高，但受制于资源利用率水平较低，导致发展不平衡，出现了一定程度管理不当现象。图 4 和图 5 为各指数 2019-2020 年与 2018-2019 年两个时段的差值，也佐证了第二时段综合发展情况突出，但技术效率发展限制了“开源中国”社区的整体发展水平。

表 7 “开源中国”社区 2018-2020 年分时段各板块 Malmquist 指数

板块	2018-2019					2019-2020				
	effch	techch	pech	sech	tfpch	effch	techch	pech	sech	tfpch
Java	1.000	0.819	1.000	1.000	0.819	1.000	1.195	1.000	1.000	1.195
Android	1.573	1.836	1.370	1.148	2.887	0.648	1.459	0.768	0.843	0.946
PHP	1.004	0.950	1.000	1.004	0.953	1.000	1.160	1.000	1.000	1.160
MySQL	0.821	0.940	0.831	0.988	0.771	1.119	1.500	1.204	0.929	1.679
Spring	1.164	0.573	1.000	1.164	0.667	0.699	1.094	0.714	0.978	0.764
JFinal	1.000	1.509	1.000	1.000	1.509	1.000	2.863	1.000	1.000	2.863
Python	0.999	0.974	1.000	0.999	0.972	0.642	1.336	0.646	0.994	0.857
Eclipse	1.132	0.988	1.109	1.020	1.118	0.622	1.100	0.649	0.959	0.684
Linux	1.151	1.652	1.000	1.151	1.901	1.000	1.268	1.000	1.000	1.268
jQuery	1.588	0.903	1.145	1.386	1.434	0.608	1.521	1.000	0.608	0.925
Apache Tomcat	0.685	0.780	0.723	0.947	0.534	1.474	1.529	1.384	1.065	2.254
Ubuntu	1.000	1.526	1.000	1.000	1.526	0.537	1.667	1.000	0.537	0.895
CentOS	1.347	0.773	1.386	0.972	1.042	1.286	1.134	1.164	1.105	1.458
Android SDK	1.566	1.839	1.362	1.150	2.881	0.648	1.459	0.768	0.843	0.946
Apache ECharts	1.000	0.389	1.000	1.000	0.389	1.000	1.696	1.000	1.000	1.696
均值	1.105	1.006	1.046	1.057	1.112	0.846	1.420	0.931	0.908	1.201

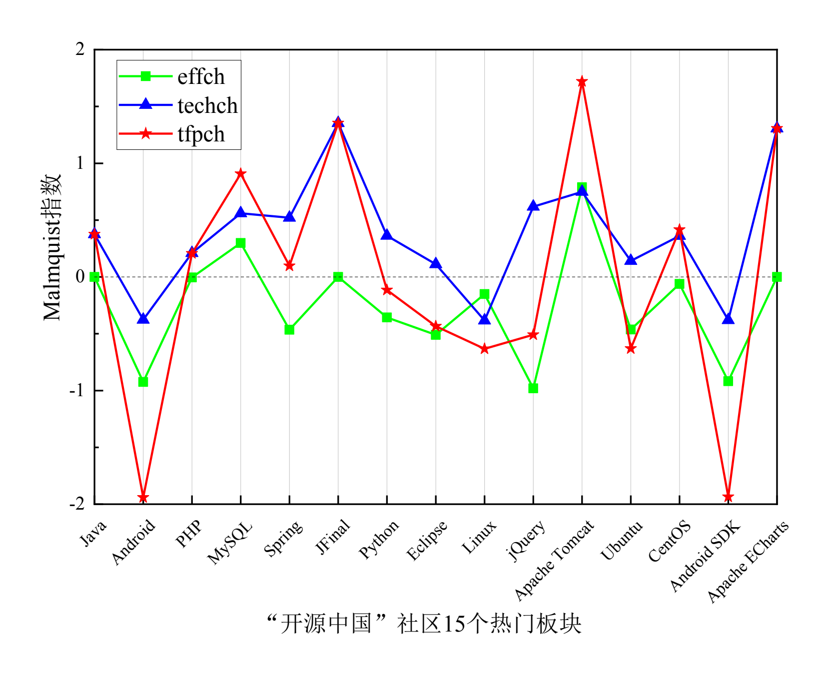


图 4 “开源中国”社区各板块两时段间全要素生产率分解变动趋势

知识管理论坛，2022（2）:101-115
DOI: 10.13266/j.issn.2095-5472.2022.009

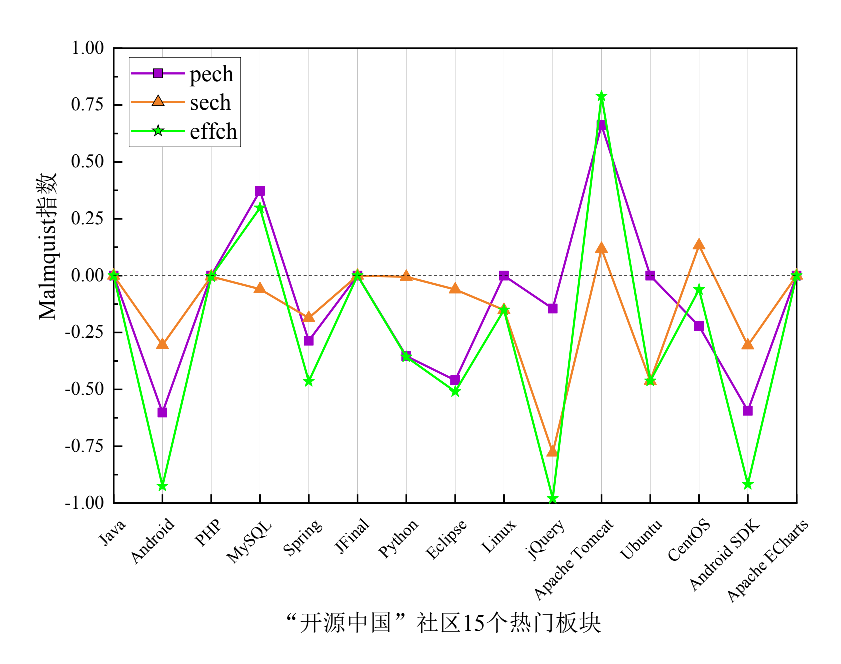


图 5 “开源中国”社区各板块两时段间技术效率分解变动趋势

5 结论与建议

本文在已有研究的基础上，基于 SBM-Malmquist 对技术问答社区知识交流效率进行测度评价，结果表明：2018-2020 年间“开源中国”社区 15 个热门板块平均知识交流效率良好，但板块间的差距较大，每年仅有 3 个板块达到 DEA 有效，并且板块知识交流效率均值呈现“先上升后下降”的波动变化，以 2020 年最为明显，说明如疫情等重大事件可能会造成社区知识交流出现明显下降；无效单元中仅存在 Android 板块在 2019 年规模效率有效，说明板块资源配置水平参差不齐，投入资源并未得到优化配置，投入冗余而导致资源得不到充分利用；全要素生产率处于上升趋势，平均涨幅 15.60%，技术效率变化呈现下降趋势，技术进步变化指数呈现 19.50% 的上涨幅度，且 2019-2020 年间上涨了 41.40%，充分说明技术振兴是“开源中国”社区知识交流效率上升的主要原因，但存在管理不够完善及资源利用率不够高的问题仍亟待解决。

基于上述研究结果，本文从运营管理角度出发，为提高技术问答社区知识交流效率提出以下建议：

第一，整合技术问答社区资源管理。社区管理者应根据自身定位，积极对标国外先进技术问答社区，结合自身实际情况吸纳有效改进措施。同时，合理定位用户群体，针对 Java、JFinal 和 Apache Echarts 等每年均达到 DEA 有效的板块设置相应的门槛，保证用户质量与数量的同时打造出其特有的文化氛围。另外，技术问答社区的有序运行不仅依赖于用户的发帖、投票和浏览，社区管理者应主动引导，例如定期发起时效性强的热点内容吸引用户参与并讨论，同时规避由于技术问答社区的不断发展而引发的信息激增现象，这会导致内容重复、低相关性和低质量的信息大量堆积，从而影响用户体验效果。

第二，优化技术问答社区资源配置。可以采用集约式管理，针对资源配置水平较低的板块从量的提升逐步过渡到质的把控，有针对性地根据内容对帖子进行筛选，例如管理员可通过选取精华帖子并邀请具有较高活跃度和荣誉值的用户进行回帖，同时为回帖和发帖质量等

级相对较高的用户提供一定的激励，如社区虚拟金币等。同时，优化投票机制，同一回帖者下的再回帖者和发帖者均设置投票功能，并对得票数量高的用户给予一定的物质和精神奖励，从而不断调整社区的内部激励机制。除此之外，社区可以在主页或功能栏为新用户设置实操性较强案例的教程，提高用户信息素养，提升用户黏性，从而优化技术问答社区单位投入的所得产出量。

第三，推动技术问答社区技术创新。可以设置同类问题的推送功能，例如根据用户的发帖、浏览与回帖等情况锁定用户所感兴趣问题的搜索进程和轨迹，通过对用户深度挖掘并聚类构建用户画像，可以更精准地进行同类问题的推送并及时推荐可参考的相关评论，同时设置内部监管，屏蔽不良信息，使用户尽可能多地获得有用信息。另外，人工智能技术的进步可以促进技术问答社区人机对话模式的形成，匹配数据库将同类问题的答案自动回复给用户，并可以根据用户所提关键词引导用户进入相关领域或展现类似知识图谱的推荐功能，激发用户进一步思考并提升知识交流效率。

本文所选取的研究对象和方法具有一定的创新性，但也存在着一定的不足和局限，在指标体系的构建中，根据技术问答社区自身特征加入了“帖子投票数量”这一新指标，但只从微观层面对知识交流效率进行了测度，对于宏观层面上影响知识交流效率的因素有待进一步研究，因此在未来持续跟进研究中，可将数据包络分析法与其他方法相结合，在测度知识交流效率的同时，进一步分析知识交流效率的影响因素，从而对技术问答社区提出更深层次改善建议。

参考文献：

[1] 彭红彬 , 王军 . 虚拟社区中知识交流的特点分析——基于 CSDN 技术论坛的实证研究 [J]. 现代图书情报技术 , 2009(4): 44-49.

[2] 李胜利 , 钟滢 . 中外技术问答社区的实证对比研究与启示——以 CSDN 和 Stack Overflow 为例 [J]. 情报学报 , 2020, 39(9): 989-1000.

[3] 蒋竞 , 吕江枫 , 张莉 . 中文软件问答社区主题分析研究 [J]. 软件学报 , 2020, 31(4): 1143-1161.

[4] 刘雨农 , 刘敏榕 . 虚拟知识社区的社会网络结构及影响因素——以知乎网为例 [J]. 图书情报工作 , 2018, 62(4): 89-96.

[5] LIU Z, JANSEN B J. Questioner or question: predicting the response rate in social question and answering on Sina Weibo[J]. Information processing & management, 2018, 54(2): 159-174.

[6] JIN J H, LI Y J, ZHONG X J, et al. Why users contribute knowledge to online communities: an empirical study of an online social Q&A community[J]. Information & management, 2015, 52(7): 840-849.

[7] 沈洪洲 , 史俊鹏 , 马巧慧 . 社会化问答社区回答内容质量影响特征研究——以“知乎”为例 [J]. 情报杂志 , 2020, 39(10): 169-175, 202.

[8] 王伟 , 冀宇强 , 王洪伟 , 等 . 中文问答社区答案质量的评价研究 : 以知乎为例 [J]. 图书情报工作 , 2017, 61(22): 36-44.

[9] 李丹 . 中美网络问答社区的对比研究——以 Quora 和知乎为例 [J]. 青年记者 , 2014(26): 19-20.

[10] 万莉 . 学术虚拟社区知识交流效率测度研究 [J]. 情报杂志 , 2015, 34(9): 170-173.

[11] 闫安 , 李添秀 . 虚拟社区中知识获取方式对获取结果的影响研究——以知乎社区为例 [J]. 图书馆理论与实践 , 2020(1): 65-71, 87.

[12] 庞建刚 , 吴佳玲 . 基于 SFA 方法的虚拟学术社区知识交流效率研究 [J]. 情报科学 , 2018, 36(5): 104-109.

[13] 宗乾进 , 吕鑫 , 袁勤俭 , 等 . 学术博客的知识交流效果评价研究 [J]. 情报科学 , 2014, 32(12): 72-76.

[14] 晋升 . 基于 DEA 方法的学术虚拟社区知识交流效率研究 [D]. 郑州：郑州大学 , 2019.

[15] 杨瑞仙 , 黄书瑞 , 王元锋 . 基于三阶段 DEA 模型的在线健康社区知识交流效率评价研究 [J]. 情报理论与实践 , 2020, 43(10): 122-129.

[16] 胡德华 , 张又月 , 罗爱静 . 基于遗传投影寻踪算法的学术虚拟社区知识交流效率研究 [J]. 图书馆论坛 , 2019, 39(4): 67-73, 83.

[17] 吴佳玲 . 虚拟学术社区知识交流效率研究 [D]. 绵阳：西南科技大学 , 2019.

[18] 袁永旭 , 张亚飞 , 马瑞敏 , 等 . 基于 SBM-Tobit 模型的在线健康社区知识交流效率研究 [J]. 情报科学 , 2021, 39(5): 106-114.

[19] 魏权龄 . 数据包络分析 [M]. 北京 : 科学出版社 , 2004.

[20] 陈晓红 , 易国栋 , 刘翔 . 基于三阶段 SBM-DEA 模型

的中国区域碳排放效率研究 [J]. 运筹与管理 , 2017, 26(3): 115-122.
[21] 万莉 , 程慧平 . 管理科学部重要期刊知识交流效率评价——基于 Super-SBM 与 SFA 模型的实证研究 [J]. 现代情报 , 2017, 37(11): 69-73.
[22] TONE K. A slacks-based measure of efficiency in data envelopment analysis[J]. European journal of operational research, 2001, 130(3): 498-509.
[23] 孙红蕾 , 马岩 , 郑建明 . 城市信息基础设施效率测评研究 [J]. 图书馆论坛 , 2017, 37(5): 1-9.
[24] 杨瑞仙 , 张广轶 . 学术虚拟社区知识交流过程与机理研究 [J]. 现代情报 , 2020, 40(10): 52-61.
[25] 王俊俊 . 基于 SBM-Malmquist 的广西旅游生态效率研究 [D]. 桂林：桂林理工大学 , 2020.
[26] 方叶林 . 中国省域旅游业效率及演化机理研究 [D]. 南京：南京师范大学 , 2014.
[27] 洪图 , 李飚 . 我国农业类上市公司研发效率测度与划分——基于 SBM-Malmquist 模型和企业生命周期的实证分析 [J]. 经济问题探索 , 2020(9): 65-77.
[28] 陈明红 , 谢晓惠 . 网络信息资源生态化配置效率评价研究 [J]. 情报理论与实践 , 2020, 43(10): 81-87.

作者贡献说明：
丁　楠：设计和改进研究方案，修改论文；
曹玮倬：提出研究选题，搜集与分析数据，撰写与修改论文；
相蒙蒙：参与研究方案、研究思路设计。

Research on Knowledge Exchange Efficiency Measurement of Technical Question-Answering Community Based on SBM-Malmquist

Ding Nan Cao Weizhuo Xiang Mengmeng
School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun 130117

Abstract: [Purpose/Significance] The current research on knowledge exchange in technical Q&A communities is not perfect. This article uses data envelopment analysis method to study the efficiency of knowledge exchange in technical Q&A communities, and finds the differences and changing laws between different sectors to provide references for improving community management and improving the efficiency of knowledge exchange. **[Method/Process]** Based on the construction of the knowledge exchange efficiency evaluation index system, 15 popular sections of the “OSCHINA” community were used as the research samples, and the static and dynamic efficiency values of the knowledge exchange are calculated and analyzed deeply based on the SBM model and the Malmquist index. **[Result/Conclusion]** The average knowledge exchange efficiency among the 15 popular sections of the “OSCHINA” community is good, but the gap between the sections is large, and the overall knowledge exchange efficiency is on the rise trend, mainly because the technological progress change index increases 19.50%. The increasing rate indicates that technological revitalization is the main reason for the increase in the efficiency of knowledge exchange in the “OSCHINA” community, but a certain degree of mismanagement has occurred and the utilization of resources is low.

Keywords: technical question-answering community SBM model Malmquist index knowledge exchange efficiency OSCHINA